Osteoarthritis and Cartilage (2008) **16,** 631–634

doi:10.1016/j.joca.2007.10.010

Osteoarthritis and Cartilage

International Cartilage Repair Society

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Brief report
Development of a predictive model for estimating the probability of treatment success one year after total shoulder replacement – cohort study[1]

B. R. Simmen M.D., Chairman†, L. M. Bachmann M.D., Ph.D., Deputy Director‡, S. Drerup, Research Fellow†, H.-K. Schwyzer M.D., Consultant Orthopedic Surgeon, Deputy Chairman†, A. Burkhart M.D., Research Fellow‡ and J. Goldhahn M.D., Senior Research Fellow†§*

† *Orthopedics Department, Upper Extremity Unit, Schulthess Clinic, Lengghalde 2, 8008 Zürich, Switzerland*
‡ *Horten Centre for Patient-Oriented Research, University of Zurich, Bolleystrasse 40, Postfach Nord, 8091 Zurich, Switzerland*
§ *ETH Zurich Institute of Biomechanics, 8093 Zürich, Switzerland*

Summary

Objective: To Estimate the probability of treatment success 1 year after a total shoulder arthroplasty by developing a model based on preoperative clinical factors.

Method: Between June 2003 and December 2006, 140 patients undergoing shoulder operations were assessed for age, gender, current rheumatoid arthritis, Short Form (SF) 36 physical and mental sum scores, previous shoulder operations, the Disabilities of Arm, Shoulder and Hand (DASH) symptom and function scores, the Shoulder Pain and Disability Index (SPADI), and insurance status. One year after the operation a Constant score of 80 or more out of 100 indicated successful treatment. Patient variables were analyzed with a logistic regression model augmented in a stepwise manner and bootstrapped 100 times. Variables selected at least 33 times were incorporated into a final model and the Area under the Receiver Operating Characteristics Curve (aROC) was calculated.

Results: There were 47/140 (33.6%) successful treatments. The probability of success was reduced in patients with previous shoulder operations (Odds Ratio [O.R.] 0.17, 95% Confidence Interval (95%CI) 0.04–0.85; $P = 0.03$) and older than 75 years (O.R. 0.21, 95%CI 0.05–0.77; $P = 0.02$). The probability of success increased in patients with a higher SF 36 mental sum score (O.R. 1.03, 95%CI 0.96–1.09, $P = 0.42$) and a higher DASH function score (O.R. 1.05, 95%CI 1.02–1.07, $P = 0.001$). The aROC was 0.79 (0.70–0.88) indicating that the model has a high predictive capacity.

Conclusion: Once validated this model based on four preoperative clinical factors offers a prediction of whether a patient will respond to treatment 1 year after total shoulder arthroplasty.

Key words: Surgery, Shoulder replacement arthroplasty, Prognosis, Treatment outcome, Models, biological.

Introduction

Estimating treatment success for patients undergoing shoulder arthroplasty is difficult for surgeons due to the lack of predictive instruments. More broadly, the general field of orthopedics lacks statistical models for predicting the probability of treatment outcomes. We are aware of only a few published studies, including: a study focused on assessing treatment failure after total hip replacement in a medicare population of the United States[1], a follow-up study focused on determining predictors of non-recovery for orthopedic patients after minor traffic accidents[2], and a paper focused on identifying factors that predict complications after total knee replacement[3].

Predicting the outcome of shoulder surgery for individual patients is difficult because shoulder stability and movement depend on the complex coordination of five functional joints[4]. Shoulder disabilities directly translate into dysfunctional performance in everyday life. Since prognosis between patients varies considerably, orthopedic surgeons face a major challenge in properly informing patients regarding the risks and benefits of corrective interventions. Any intervention, whether conservative or surgical, poses the risk of poor long-term outcomes. While a failure of non-surgical interventions might delay effective measures, surgery represents a substantial traumatic stress. Selecting the most beneficial treatment for each patient requires a tool that provides individualized prognostic information based on the facts available at the patient–doctor encounter.

[1]Funding: this work was supported by the Georg and Bertha Schwyzer-Winiker Foundation, Zurich, Switzerland. Dr Bachmann's work was supported by the Swiss National Science Foundation (Grants 3233B0-103182 and 3200B0-103183).
*Address correspondence and reprint requests to: Dr Jörg Goldhahn, M.D., Senior Research Fellow, Orthopedics Department, Upper Extremity Unit, Schulthess Clinic, Lengghalde 2, 8008 Zurich, Switzerland. Tel: 41-44-255-86-54; Fax: 41-44-255-97-20; E-mail: joerg.goldhahn@kws.ch
Received 19 June 2007; revision accepted 15 October 2007.

The goal of this study was to use preoperatively available clinical parameters in developing a model for estimating the probability of treatment success 1 year after total shoulder replacement. Previous studies investigated risk factors involved in treatment failure in patients undergoing either hemi or total shoulder arthroplasty[5–7]. However, we are unaware of any study that developed a prognostic model based on preoperative clinical parameters that predicts treatment outcomes after total shoulder arthroplasty. Since osteoarthritis is one of the more common indications for shoulder arthroplasty our study population primarily consists of patients with primary or secondary osteoarthritis of the shoulder.

Methods

PATIENTS

This study was approved by the institutional ethics committee of the orthopedic sub-committee and the cantonal ethics committee of Zurich, Switzerland. The cohort consisted of a consecutive series of 140 consenting eligible patients, who underwent total shoulder replacement therapy at the Orthopedics Department, Upper Extremity Unit, Schulthess Clinic, Zurich, Switzerland. Patients were enrolled in the study between June 1, 2003 and December 31, 2006. All patients received the same PROMOS™ prosthesis.

PREDICTOR VARIABLES

A set of candidate indicator variables were defined *a priori*, including the following: age above/below 75 years, gender, current indication of rheumatoid arthritis, the Short Form (SF) 36 physician and mental sum scores (continuous scale), any previous shoulder operations, the Disabilities of the Arm, Shoulder and Hand (DASH) symptom score (continuous scale from 0 = worst to 100 = best), the Shoulder Pain and Disability Index (SPADI) (continuous scale from 0 = worst to 100 = best), and insurance status (private insurance/else).

The DASH Score is a 30-item, self-report questionnaire which was designed to measure physical function and symptoms in people with any of several musculoskeletal disorders of the upper limb[8]. The SPADI is a self-report questionnaire developed to measure the pain and disability associated with shoulder pathology. The SPADI consists of 13 items in two subscales: pain (five items) and disability (eight items); items were presented in a visual analog format[9]. We used the translated and validated German versions[10,11] of both questionnaires. We administered the questionnaires 1 day prior to the operation for all patients. For validity and reliability of these questionnaires we refer to our previous publication[12].

OUTCOMES

All patients were evaluated with the Constant score[13] in a 1-year follow-up exam. This scoring system consists of four variables that are used to assess the function of the shoulder. The right and left shoulders are assessed separately. Two subjective variables pain and activities of daily living (ADL) (sleep, work, and recreation/sport) are assigned a maximum of 35 points. Two objective variables, range of motion and strength, are assigned a maximum of 65 points. A total score of 100 points indicates perfect function. For the purpose of this analysis, a Constant score value of 80 or more was considered a treatment success. A score of 80 corresponds to the following parameter assessments in the operated shoulder: no pain, 75% capability for work, 75% capability for sports and leisure activities, no problems during sleep, full arm movement above the head, with anteversion and abduction between 121 and 150°, ability to place the hand above and behind the head with elbow ventral, but not with elbow lateral, unable to fully extend the hand above the head, ability to rotate arm internally until the hand reaches L3, and 9 kg abduction strength measured with a spring balance. We did not adjust score values for age and gender. All assessments were evaluated by the same personnel using a standardized assessment protocol to minimize variability.

ANALYSIS

Patient selection

Model derivation and validation were performed with different groups of patients. Patients were assigned to groups by a random drawing; approximately 2/3 of the patients ($n = 100$) formed the group used to develop the model (derivation group), and the 1/3 of the patients ($n = 40$) formed the group used to validate the model (validation group).

Model construction

A logistic regression model was augmented in a stepwise manner (P-value for entry = 0.05)[14]. We bootstrapped this procedure 100 times and examined the selection frequency for each variable[15]. The bootstrap is an elegant method for reducing the probability that a variable is selected due to idiosyncrasies of the data sample. This procedure randomly resamples a set of data from a given number of patients and repeats the logistic regression using a stepwise augmentation. A high selection frequency indicates that the selection of a variable is not highly dependent on the distribution of the original data set. We arbitrarily defined a selection frequency of ≥33% as a threshold for identifying variables to be used in building the model. A list of selection frequencies for all the assessed variables is available in Table II.

Model validation

The model was validated on patients in the validation group and its discrimination and calibration were assessed using the Brier Score[16]. The predictive capacity of the final model was quantified by calculating the Area under the Receiver Operating Characteristics Curve (aROC).

The coefficients in the logistic regression function (The Simmen Risk Calculator) were derived using all 140 patients. Odds ratios (O.Rs.) and 95% Confidence Intervals (95%CI) were calculated. No overfitting was detected, thus no adjustments to the coefficients of the final model were necessary[17]. Analysis was performed using the Stata 9.2 statistical software package (StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA).

Results

PARTICIPANTS

Between June 1, 2003 and December 31, 2006 140 patients underwent total shoulder arthroplasty and had a 1-year follow-up. Table I shows the distribution of all registered data. The total number of treatment successes was 47/140 (33.6%). The treatment success rates were 31% and 40% for the derivation and validation patient groups, respectively.

VARIABLES SELECTED WITH THE MODEL

Four variables were selected as good prediction factors based on their high selection frequencies: DASH function

Table I
Mean values (±SD) or percentages where appropriate for clinical characteristics of the study population

Prognostic variables	Constant score ≥ 80 ($n = 47$)	Constant score < 80 ($n = 93$)
*Age over 75 years**	5 (11%)	31 (33%)
Male gender*	20 (43%)	75 (81%)
Rheumatoid arthritis*	3 (6%)	12 (13%)
Primary osteoarthritis of the shoulder	39 (83%)	55 (59%)
Secondary osteoarthritis of the shoulder	5 (11%)	25 (27%)
Fresh fracture	0 (0%)	1 (1%)
SF 36 physical sum score*	40.7 (9.2)	35.6 (9.0)
*SF 36 mental sum score**	57.8 (6.2)	54.5 (10.0)
*Previous shoulder operation**	2 (4%)	17 (18%)
DASH symptom score*	55.9 (15.8)	49.8 (16.0)
*DASH function score**	64.2 (15.7)	52.0 (18.8)
SPADI ≥ 40*	26 (55%)	28 (30%)
Private insurance status*	35 (74%)	57 (61%)

Variables in italics constituted the final model.
*Variables assessed to be included in the model.

score (69 times), the SF 36 mental sum score (45 times), at least one previous shoulder operation (43 times), and age over 75 years (41 times). The remaining variables and the corresponding frequencies are available in Table II.

PROGNOSTIC MODEL PREDICTIONS

Using data from all the patients ($n = 140$), the prognostic model predicted a low probability of treatment success in patients with previous shoulder operations (O. R. 0.17, 95%CI 0.04−0.85, $P = 0.03$) and older than 75 years (O.R. 0.21, 95%CI 0.05−0.77, $P = 0.02$). The model predicted a high probability of treatment success in patients with a high SF 36 mental sum score (O.R. 1.03, 95%CI 0.96−1.09, $P = 0.42$) and a high DASH function score (O.R. 1.05, 95%CI 1.02−1.07, $P = 0.001$). The aROC was 0.79 (95%CI 0.70−0.88) indicating that the model has a high predictive capacity.

SCORING FUNCTION AND CORRESPONDING PROGNOSTIC PROBABILITY FUNCTION (PF)

The scoring function S and the corresponding logistic PF (The Simmen Risk Calculator), $PF = 1/[1 + \exp(-S)]$, where $S = -4.37 - 1.75 * \text{previous shoulder operations} - 1.58 * \text{age over 75 years} + 0.025 * \text{SF 36 mental score} + 0.044 * \text{DASH function score}$.

For example, the estimated probability of a treatment success in over 75 years old patient with an SF 36 mental score value of 45, a DASH score of 50 and previous shoulder operations would be $1/[1 + \exp-(-4.375)] = 1.2\%$.

On the other extreme, a 60-year-old patient without previous shoulder operations and an SF-mental score value of 65 and a DASH score of 75 has an estimated 1-year probability of treatment success of 71.5%.

Discussion

This study reports the development of a prognostic model based on four preoperative clinical factors that predicts treatment success 1 year after total shoulder replacement arthroplasty in a population that primarily consists of patients with primary or secondary osteoarthritis. To our knowledge this is the first study to offer a tool for assessing the probability of a successful shoulder replacement operation. We believe that this model will help physicians to discuss the pros and cons of a total shoulder replacement arthroplasty with their patients.

Table II
Frequency with which a single variable was selected in the bootstrap stepwise procedure. A higher number indicates that the selection of a variable is less dependent on the distribution of the original data set

Indicator	Number of entries
DASH function score	69
SF 36 mental score	45
Previous shoulder operation	43
Age $\geq$ 75 years	41
SPADI $>$ 40 points	20
DASH symptom score	12
Insurance status	6
SF 36 physical score	4
Rheumatoid arthritis	0

While various authors have studied the association of preoperative clinical findings on treatment outcomes for different joints[3,5–7,18,19], only two[3,18] used multivariate analysis, and only one[3] published a prognostic model. SooHoo and colleagues used a model to predict complication rates in patients undergoing total knee replacement[3]. They showed that age, gender, race/ethnicity, comorbidity, insurance type, and hospital volume were independent predictors for complications.

In contrast to the SooHoo *et al.* report[3], our study showed a weak association between insurance status and outcome. This suggests that though patients with private insurance may arguably choose better surgeons and receive better postoperative care and rehabilitation, they do not necessarily attain better functional results. However, further research is needed to clarify this finding. Our findings partly disagree with those published by Matsen and colleagues[7], who found that the strongest correlates with postoperative shoulder function included male gender, preoperative physical function, social function, mental health, and preoperative shoulder function. However, these results are not directly comparable to our study since study conditions were different: Matsen *et al.* had a mean follow-up of 3.4 years, used other outcome parameters, and applied other statistical methods.

Interestingly, the SPADI was not chosen and the DASH was chosen as an indicator variable in our study. The SPADI focuses on the shoulder, and differs from the DASH questionnaire which focuses on the upper extremities and is not specific for the shoulder. Both questionnaires measure pain and physical functions, but the DASH includes emotional and social domains. This finding together with our finding that the SF 36 mental summary score, but not the SF 36 physical summary, is a prognostic variable suggests that emotional and social domains may be important in predicting successful outcomes after shoulder arthroplasty. We speculate that these emotional and social domains can affect one's abilities to cope with pain and participate in rehabilitation. These factors can influence the degree of functional improvement attained with rehabilitation.

The strengths of our model are its high biological plausibility and simplicity. Once confirmed it can be transferred easily to a handheld computer for immediate assessment of the probability of treatment success for an individual patient. This facilitates the surgeon's decision of whether to provide further details about the operation or discuss non-surgical alternatives.

The weaknesses of this study are its limited size and its relatively short follow-up time. Although 140 consecutive patients undergoing total shoulder arthroplasty are considered a large cohort, a larger cohort would have facilitated extensive split sample validations. Moreover, the insufficient cohort size makes it difficult to evaluate the potential risks of diagnoses other than a current rheumatoid arthritis, such as posttraumatic osteoarthritis or fracture[20]. Nevertheless, because the model needed no adjustments (shrinkage) for overfitting, we are optimistic that the coefficients we reported will be confirmed in a properly sized validation cohort. Furthermore, future investigators might choose to pool their coefficients with those presented here in order to derive a "meta-analytic" model containing more complete derivation and validation data. In addition, the generalizability of the model should be examined by applying it to other racial and ethnic groups. The relatively short follow-up time of 1 year was a limitation because we might have missed some cases of late recovery or deterioration. Lastly, we used a high critical Constant score value to define treatment success. However, there were logical reasons for using this threshold. The value of 80 for the Constant score corresponds to a shoulder joint with no pain and adequate function. Based on the perspective of the patient, pain free movement is regarded as prerequisite, while slight deficits in range of motion, strength and subsequent ADL are acceptable. Though this threshold may be ambitious it matches the perception of a successful operation held by experienced orthopedic surgeons and, more importantly, patients. Other unconventional methods are the use the SF 36 mental score and the DASH function score. It is known that mental status and patient-perceived function interact in the sense that they reflect a patient's perception independent of actual function. Currently, one or both questionnaires might not be completed routinely in all orthopedic clinics. However, since both rely on self-reporting, this information can be gained without lengthening consultation time. At our clinic all patients who are evaluated for total shoulder arthroplasty receive these questionnaires and a recent survey among patients revealed that filling the forms was unproblematic.

We believe that there is a clear need for more carefully developed and validated prognostic models. In orthopedic surgery in particular, where a substantial proportion of interventions aim at improving quality of life rather than survival, efficient tools for estimating the likelihood of treatment success would be of extraordinarily high value for physicians and patients alike. The big advantage of prognostic models is that they express health risks explicitly. In our example, the model allows discriminating between those patients who are very likely to have good to excellent postoperative function and those who are unlikely to benefit from the intervention.

Future studies may validate our model and further quantify the patient definition of treatment success and determine more precise preoperative factors that predict meaningful improvement based on those results. Studies may also explore whether shortened versions of the questionnaires like the SF 8 and the DASH would provide adequate prognostic information. In addition, further studies could expand the follow-up period to 2 years in order to capture more long-term risk factors.

In conclusion, if confirmed, this simple model based on four preoperative clinical factors allows assessing treatment success 1 year after total shoulder arthroplasty.

Conflict of interest

All authors confirm that they had no financial and personal relationships with other people or organizations that could inappropriately influence (bias) their work.

Acknowledgment

We would like to thank Dr Ulrike Held for statistical assistance and supervision.

References

1. Mahomed NN, Barrett JA, Katz JN, Phillips CB, Losina E, Lew RA, *et al.* Rates and outcomes of primary and revision total hip replacement in the United States medicare population. J Bone Joint Surg Am 2003; 85-A:27–32.
2. Ottosson C, Nyren O, Johansson SE, Ponzer S. Outcome after minor traffic accidents: a follow-up study of orthopedic patients in an inner-city area emergency room. J Trauma 2005;58:553–60.
3. SooHoo NF, Lieberman JR, Ko CY, Zingmond DS. Factors predicting complication rates following total knee replacement. J Bone Joint Surg Am 2006;88:480–5.
4. Mc Beth J, Macfalane GJ. The prevalence of regional and widespread musculoskeletal pain symptoms. In: Linton SJ, Ed. New Avenues for the Prevention of Chronic Musculoskeletal Pain And Disability. Amsterdam: Elsevier 2002:7–22.
5. Hettrich CM, Weldon E 3rd, Boorman RS, Parsons IM 4th, Matsen FA 3rd. Preoperative factors associated with improvements in shoulder function after humeral hemiarthroplasty. J Bone Joint Surg Am 2004;86-A:1446–51.
6. Iannotti JP, Norris TR. Influence of preoperative factors on outcome of shoulder arthroplasty for glenohumeral osteoarthritis. J Bone Joint Surg Am 2003;85-A:251–8.
7. Matsen FA 3rd, Antoniou J, Rozencwaig R, Campbell B, Smith KL. Correlates with comfort and function after total shoulder arthroplasty for degenerative joint disease. J Shoulder Elbow Surg 2000;9: 465–9.
8. Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the disabilities of the arm, shoulder and hand outcome measure in different regions of the upper extremity. J Hand Ther 2001; 14:128–46.
9. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. Arthritis Care Res 1991;4:143–9.
10. Angst F, Goldhahn J, Pap G, Mannion AF, Roach KE, Siebertz D, *et al.* Cross-cultural adaptation, reliability and validity of the German Shoulder Pain and Disability Index (SPADI). Rheumatology (Oxford) 2007; 46:87–92.
11. Offenbaecher M, Ewert T, Sangha O, Stucki G. Validation of a German version of the disabilities of arm, shoulder, and hand questionnaire (DASH-G). J Rheumatol 2002;29:401–2.
12. Angst F, Pap G, Mannion AF, Herren DB, Aeschlimann A, Schwyzer HK, *et al.* Comprehensive assessment of clinical outcome and quality of life after total shoulder arthroplasty: usefulness and validity of subjective outcome measures. Arthritis Rheum 2004; 51:819–28.
13. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clin Orthop Relat Res 1987;160–4.
14. Kleinbaum DG, Klein M. Logistic Regression a Self-Learning Text. 2nd edn. Springer 2002.
15. Efron B, Tibshirani RJIn: An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC 1998; vol. 57.
16. Spiegelhalter DJ. Probabilistic prediction in patient management and clinical trials. Stat Med 1986;5:421–33.
17. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15: 361–87.
18. Escobar A, Quintana JM, Bilbao A, Azkarate J, Guenaga JI, Arenaza JC, *et al.* Effect of patient characteristics on reported outcomes after total knee replacement. Rheumatology (Oxford) 2007; 46:112–9.
19. Long MJ, McQueen DA, Bangalore VG, Schurman JR 2nd. Using self-assessed health to predict patient outcomes after total knee replacement. Clin Orthop Relat Res 2005;189–92.
20. Boileau P, Sinnerton RJ, Chuinard C, Walch G. Arthroplasty of the shoulder. J Bone Joint Surg Br 2006;88:562–75.